

In the Claims

1–8. (Canceled)

9. (Currently amended) A method for providing segmentation of an input stream in a character-based language comprising:

creating a plurality of segments from at least two tokens in the input stream based upon lexical information and lexical functions for the character-based language, the lexical functions comprising computer-executable routines generated by a grammar programming language compiler; and

generating a connection graph using the plurality of segments.

10. (Original) The method of claim 9 further comprising compiling lexical grammar rules to generate the lexical functions, the lexical grammar rules being written in a grammar programming language.

11. (Original) The method of claim 10 wherein the lexical grammar rules define connectivity relation of tokens.

12. (Original) The method of claim 9 further comprising assigning at least one part of speech tag to at least one segment using a lexical dictionary.

13. (Original) The method of claim 12 further comprising:

defining a plurality of paths in the connection graph based upon part of speech tags and the segments;

assigning a cost to each of the plurality of paths; and

determining at least one best path based upon a corresponding cost to generate an output graph.

14-21. (Canceled)

22. (Currently amended) An apparatus for providing segmentation of an input stream in a character-based language, comprising:

means for creating a plurality of segments from at least two tokens in the input stream based upon lexical information and lexical functions for the character-based language, the lexical functions comprising computer-executable routines generated by a grammar programming language compiler; and

means for generating a connection graph using the plurality of segments.

23. (Original) The apparatus of claim 22 further comprising means for compiling lexical grammar rules to generate the lexical functions, the lexical grammar rules being written in a grammar programming language.

24. (Original) The apparatus of claim 23 wherein the lexical grammar rules define connectivity relation of tokens.

25. (Original) The apparatus of claim 22 further comprising means for assigning at least one part of speech tag to at least one segment using a lexical dictionary.

26. (Original) The apparatus of claim 25 further comprising:

means for defining a plurality of paths in the connection graph based upon part of speech tags and the segments;

means for assigning a cost to each of the plurality of paths; and

means for determining at least one best path based upon a corresponding cost to generate an output graph.

27-34. (Canceled)

35. (Currently amended) An apparatus for providing segmentation of an input stream in a character-based language, comprising:

a segmentation engine for creating a plurality of segments from at least two tokens in the input stream based upon lexical information and lexical functions for the character-

based language, the lexical functions comprising computer-executable routines generated by a grammar programming language compiler; and

a graph generator for generating a connection graph using the plurality of segments.

36-38. (Canceled)

39. (Original) The apparatus of claim 38 further comprising:

a path designator for defining a plurality of paths in the connection graph based upon part of speech tags and the segments;

a cost assignor for assigning a cost to each of the plurality of paths; and

a path calculator for determining at least one best path based upon a corresponding cost to generate an output graph.

40. (Canceled)

41. (Currently amended) A system for providing segmentation of an input stream in a character-based language, comprising:

a processor;
an input coupled to the processor, the input capable of receiving an input stream having at least two tokens, the processor configured to create a plurality of segments from the at least two tokens based upon lexical information and lexical functions for the character-based language, the lexical functions comprising computer-executable routines generated by a grammar programming language compiler, and generate a connection graph using the plurality of segments; and

an output coupled to the processor, the output capable of providing segmentation of the input stream.

42-46. (Canceled)

47. (Currently amended) A computer readable storage medium comprising instructions, which when executed on a processor, perform a method for providing segmentation of an input stream in a character-based language, comprising:

creating a plurality of segments from the at least two tokens in the input stream based upon lexical information and lexical functions for the character-based language, the lexical functions comprising computer-executable routines generated by a grammar programming language compiler; and

generating a connection graph using the plurality of segments.

48. (Currently amended) The computer readable storage medium of claim 47 further comprising compiling the lexical grammar rules to generate lexical functions, the lexical grammar rules being written in a grammar programming language.